IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 19 May 2023, accepted 8 June 2023, date of publication 13 June 2023, date of current version 20 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3285798

APPLIED RESEARCH

Occluded Person Re-Identification by Multi-Granularity Generation Adversarial Network

YANQI WANG[1], YANGUO SUN[2],ZHENPING LAN[1], FENGXUE SUN[1], NIANCHAO ZHANG[1], AND YURU WANG[1]

[1]Electronic Information Department, Dalian Polytechnic University, Dalian 116034, China
[2]Information Center of the Second Hospital of Dalian Medical University, Dalian 116038, China

Corresponding authors: Yanguo Sun (563083592@qq.com) and Zhenping Lan (8091811@qq.com)

This work was supported in part by the Liaoning Provincial Department of Education Project under Grant LJKZ0515 and Grant LJKFZ20220206, and in part by the Dalian Science and Technology Innovation Project under Grant 2019J13SN102.

ABSTRACT In order to address the problem that the detailed features of pedestrians are not prominent and the pedestrian pictures are obscured in unique environments in the process of person re-recognition, we propose a person re-recognition method with a multi-grain size generative adversarial network. Firstly, we use the generative adversarial network to recover the occluded pedestrian pictures; secondly, we improve the traditional multi-granularity network by adding an Efficient Channel Attention for Deep Convolutional Neural Networks (ECA-Net) on the coarse-grained branch to focus on the feature information in the pedestrian pictures and use the High-Resolution Net (HRNet) for pose estimation on the fine-grained branch to divide the pedestrian pictures into nine parts, to enhance the network's learning of more detailed features of pedestrians, and thus improve the accuracy of pedestrian re-recognition learning, which in turn improves the accuracy of person re-identification.

INDEX TERMS Person re-identification, generative adversarial networks, random occlusion, attention mechanism.

I. INTRODUCTION

Person re-identification (ReID) is a class of image retrieval problem currently receiving attention from academia and many social fields, which can realize cross-device image acquisition function. When the pedestrian target needs to be retrieved, the pedestrian image library can be retrieved with image judgment technology in computer recognition mode to determine whether the pedestrian target is in the current range, which is also the key technology of the current video surveillance intelligence, has been widely used in many fields such as criminal security investigation, and also has a broad development prospect in the unmanned supermarket, intelligent equipment development and other fields. This makes the ReID problem gradually become a research hotspot.

The associate editor coordinating the review of this manuscript and approving it for publication was Davide Patti.

Traditional ReID places more emphasis on the visual feature and similarity measure levels [1] The manual features involved during this period are Gabor features [2], Histogram of Oriented Gradient (HOG) [3], Scale Invariant Feature Transform (SIFT) [4] etc. In some works [5], [6], [7], [8], a particular pedestrian image is represented by a combination of multiple manual features. Then the similarity between features is calculated using the distance metric. However, the limitations of using traditional image extraction techniques are apparent as they cannot complete the extraction of advanced visual features. This technique is helpful for static and straightforward image discrimination but cannot adapt to image extraction of complex scenes with dynamic features.

Moreover, as the application area of deep learning continues to develop, pedestrian re-recognition is gaining more significant applications in practice. Most of the pedestrian recognition techniques built based on deep learning are based

on the direct extraction of salient pedestrian appearance features by Convolutional Neural Networks (CNN) [9], focusing on obtaining the overall characteristics of pedestrians to distinguish different pedestrians [10].

However, the effect of ReID will be affected by a series of factors, such as environmental factors, the low resolution of pedestrian images, and the occlusion of pedestrian images, so the accuracy of ReID is always maintained at a low level.

In order to better solve ReID at night, some corresponding RGB-IR Person Re-Identification [11] model has been proposed. The primary strategy of these methods is feature alignment, that is, through some network structures, The loss function is designed to map two different data into a feature space to reduce their modality gap.

However, the ReID dataset was collected in clear weather without considering that the images would be affected by bad weather such as snow, rain and fog. In these adverse environments, the visibility of the images is very low, so Kanwal et al. [12] addressed the problem of adversarial fog attack by the dark channel prior (DCP) method and used a fusion algorithm to fuse the handcrafted features with the features of the neural network to improve the effectiveness of ReID. Pang et al. [13] proposed a novel Interference Suppression Model (ISM) to cope with the effect of severe weather on ReID.

To address the above problem of low image resolution can be solved by the single image super-resolution (SISR) method. However, the model parameters and complexity of SISR are significant, so Zhu et al. [14] proposed a lightweight single image super-resolution network EMASRN image that can balance the number of parameters and performance. However, this method ignores the local features of the image during high-resolution image generation, so Zhu et al. [15] proposed two innovative mechanisms, including the cross-view block (CVB) and the spatial perception module (SPM), to fuse the information of global and local features thus improving the quality of the resolution reconstruction.

To address the above problem of low image resolution can be, solved In order to solve the pedestrian mentioned above occlusion problem, Wang et al. proposed Multiple Granularity Network (MGN) [16]. However, the effect could not be more satisfactory, so this paper proposes a multi-granularity generative adversarial network to solve the problem. Firstly, the GAN network is used to recover the occluded pedestrians, thus weakening the negative impact of the occluders on the pedestrian images. However, the effect of traditional GAN on the recovery of occluded images could be better. To improve the accuracy of pedestrian re-identification, this paper proposes to use the Aggregated Contextual Transformation GAN (AOT-GAN) [17] to recover the occluded pedestrian images.

Then the global and local features of the pedestrian are extracted and fused by a reidentification network to obtain more information about the pedestrian. However, this method could be more effective for the case of this paper, so this paper performs pose estimation to obtain more local features to improve the accuracy of ReID.

In response to the above pedestrian re-identification problem, the main contributions of this paper are highlighted as follows:

(1) This paper proposes a pedestrian re-identification method of a multi-granularity generative confrontation network, which will improve the re-identification effect of occluded pedestrians

(2) This paper replaces the traditional GAN with an AOT-GAN, which improves the network's resilience to the occluded parts of pedestrians.

(3) This paper introduces the ECA-Net module, which improves the network's ability to extract coarse-grained features.

(4) This paper uses HRNet for pose estimation to divide the fine-grained branch pedestrian pictures into nine parts so that the network can obtain more information about pedestrians, thereby improving the ReID ability of the network.

II. OVERALL NETWORK

A. OVERALL NETWORK STRUCTURE

In this paper, we design an obscured ReID strategy based on multi-grain and generative adversarial networks with the following networks: a generative adversarial network that performs the obscuration removal function and a multi-grain network that implements ReID. The overall framework can be seen in Figure 1, which includes a generator G and a discriminator D, as well as a re-identification network. The generator can restore the occluded content of the occluded image and transmits the deblocked image to the discriminator, which then processes the newly generated image and the original image. The generative adversarial network contains much information about the generated and original images. The specific process is as follows.

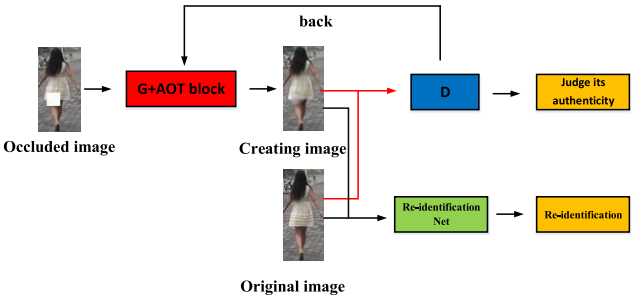

FIGURE 1. The overall framework of the network. The AOT block (as shown in Figure 4) is added to the generator G of the traditional GAN network, the image is restored by the new generator G and the discriminator D, and the restored image is passed into the re-identification network.

First, the original training image library selects complete and clear training images. The selected unobscured images are randomly masked to obtain the desired masked training images and form one-to-one pairs of masked and unmasked

images. The adversarial network is generated by training many image pairs.

Next, the trained generator is used to process the images containing random occlusions to ensure no occlu-sions in the processed images. Then, together with their original identity labels, they are transferred to the original training image set so that the desired training image set can be successfully obtained.

Finally, the improved multi-granularity network is trained with the image set to facilitate the subsequent ReID process accurately. The backbone of the improved multi-granularity network framework is the ResNet-50 network, which generates three branches in res_conv4, i.e., global, coarse-grained and fine-grained branches. The ECA attention mechanism is added after the coarse-grained branch conv5 to make the convolutional network pay more attention to the pedestrian features of the coarse-grained branch. The fine-grained network, on the other hand, improves the part of the MGN network that divides the pedestrian picture into three parts using HRNet for pose estimation into dividing the human picture into nine parts to improve the recognition effect. It is shown in Figure 2.

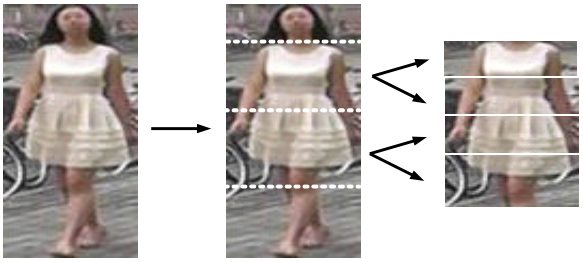

FIGURE 2. Pedestrian pictures are divided into 9 parts, these 9 parts include the head, chest, abdomen, thighs, feet, and the chest and abdomen can be subdivided into two parts: the upper part and the lower part.

B. GENERATING ADVERSARIAL NETWORK

The generative adversarial network is commonly used in artificial intelligence technology and is a technique to build a model based on deep learning technology. Generative adversarial networks generally include two components: generator and discriminator. The generator's primary function is to generate objects infinitely close to the actual image based on learning the distribution features of the real image and then uploading the generated results to the discriminator. The discriminator is a network with discriminative and classification functions, which can discriminate the images inputted by the generator and then transmit the discriminative results back to the generator to provide information for the generator's game of adversarial. The classical training process of the adversarial generative network includes the following procedure. First, the actual image and the generated image given by the generator are input into the discriminator to complete the training of the discriminator. The actual degree of the generated image will gradually improve with the training of the discriminator, and the discriminator's discriminative ability and classification ability will be enhanced; finally, it enters the convergence stage, and the discriminator no longer carries out the recognition of the actual degree of the image, which means that the generated image and the actual image have been consistent, has reached the equilibrium state. The objective function of a traditional GAN network can be expressed as:

$$\min_G \max_D V(D,G) = E_{x\sim P_{data}(x)}\left[\log D(x)\right] + E_{z\sim P_z(z)}\left[\log\left(1 - D\left(G\left(z\right)\right)\right)\right] \quad (1)$$

where G is the generator; D is the discriminator; x is the real image; Z is the image input to the generator; $G(z)$ is the image generated by the generator; $D(x)$ is the probability that the discriminator determines that x is the real image; $D(G(z))$ is the probability that the discriminator determines that the image generated by the generator is the real image; $E_{x\sim P_{data}(x)}$ and $E_{z\sim P_z(z)}$ are the expectation functions.

The results are not very satisfactory when using tra-ditional generative adversarial networks for recovering oc-cluded images, so this paper uses an AOT-GAN.

It is an enhanced type of generative adversarial network. The effect comparison is shown in Figure 3. the top 3 images on the right are those recovered by the traditional GAN network, and the bottom three images are those recovered by AOT-GAN. It can be compared that the images recovered by AOT-GAN are closer to the original images. AOT-GAN is mainly a generator that adds the AOT block to the traditional GAN. The AOT block is designed to capture information-rich long-range image contexts for contextual inference in image restoration. More details of the AOT block design are discussed below.

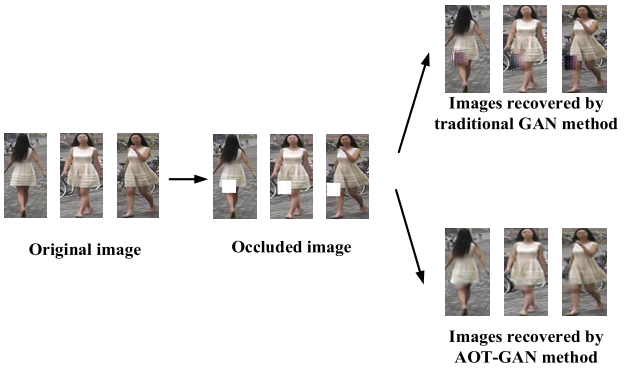

FIGURE 3. Comparison of the effect of restored pictures.

The AOT module uses three steps of splitting, converting and merging to help obtain more information [17].

Splitting: As shown in Figure 4, the AOT module splits the standard convolutional kernel into multiple sub-kernels, each with fewer output channels. For example, a kernel with 256 output channels is split into four sub-kernels, so each sub-kernel has 64 output channels.

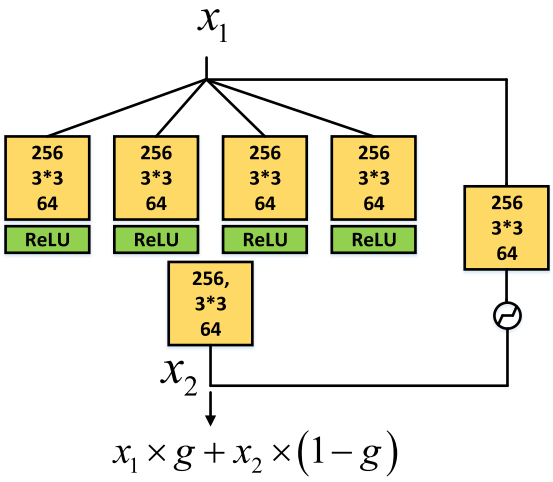

FIGURE 4. The numbers in the orange blocks in an AOT block indicate the input channels, the size of the convolution kernel and the output channels.

Transformation: Each sub-kernel performs a different transformation on the input features using different dilation rates. A more significant dilation rate allows the kernel to "see" larger areas of the input image. Nevertheless, using a lower dilation rate, the kernel focuses on local patterns in smaller fields of perception.

Merging: Contextual transformations from different sensory fields are finally aggregated in tandem, and then feature fusion is performed by standard convolution. Such a design allows the AOT module to predict each output pixel through different views.

With these three steps, the AOT module can aggregate multiple contextual transformations to enhance contextual reasoning. Due to the great success of ResNet [18]'s an excellent success, network models usually include the same residual connections in their building blocks to simplify the training of the network. However, they ignore the differences between the input pixel values inside and outside the missing regions, leading to the problem of colour discrepancies in the restored images. In order to alleviate the above problems, it is proposed to use a new gated residual connection in the building block. As shown in Figure 4, the residual connection first computes spatially varying gate values from x_1 via standard convolution and Sigmoid operations g, and then the AOT module learns the residual features by weighting with g and aggregating the input features x_1 and learning the residual features x_2, which are represented as

$$x_3 = x_1 \times g + x_2 \times (1 - g) \tag{2}$$

C. ECA-Net

Studies have demonstrated that adding attention modules to convolutional neural networks can significantly improve performance. However, the majority of current methods focus on creating attention modules that are more intricate in order to attain higher performance, which necessarily makes the model more complex. By suggesting a local cross-channel interaction method (ECA module) without dimensionality reduction and an adaptive selection of the one-dimensional convolutional kernel size to accomplish performance optimization, ECA-Net primarily adds specific enhancements to the SE-Net [19] module. Even though the module adds a few parameters, it significantly improves performance. Avoiding dimensionality reduction is crucial for learning channel attention, and effective cross-channel interaction can drastically reduce model complexity while maintaining excellent performance. As a result, the coverage of local cross-channel interactions can be established by setting the size of the one-dimensional convolutional kernel adaptively. The ECA module is chosen since it is effective and practical in light of the experiments in this study. In Figure 5, the ECA module is displayed.

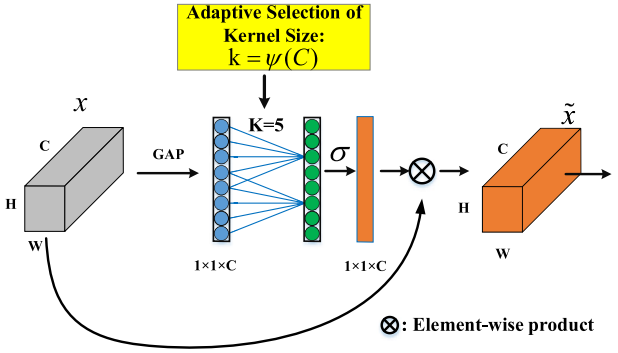

FIGURE 5. Schematic diagram of ECA module, where K = 5. Add this module to the coarse-grained branch (branch2 in Figure 7) of the entire re-identification network (as shown in Figure 7) to improve the effect of re-identification.

D. HIGH-RESOLUTION NET(HRNet)

Traditional machine learning models downsample the layers and then upsample the feature layers to recover the original layer size. However, serially connected network models such as U-Net [20] do not need to downsample and retain high-resolution features, and they cause model complexity and exponential growth of computer operations. The parallel connection of HRNet [21] can solve the above problems and complete multi-scale feature integration by repeatedly integrating feature layers of the same level and multiple levels. To a certain extent, the integration operation reduces the semantic gap in information integration, enabling the model to enhance the ability of contextual feature extraction significantly. In conclusion, HRNet can increase the accuracy of ReID, maintain more information about pedestrians with limited computational resources, and prevent the loss of details in ReID. The organization of the HRNet backbone network is depicted in Figure 6. A 3-level sub-network from high- to low-resolution is formed by combining three parallel sub-networks with various resolutions. To add low-high-resolution features, many feature fusions across sub-networks exist. Information exchanges between parallel subnetworks

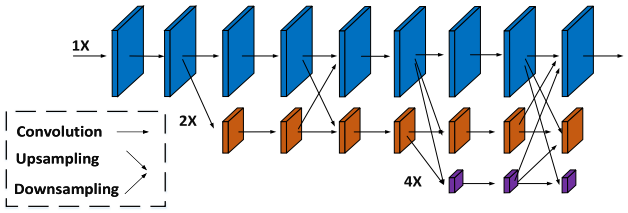

FIGURE 6. HRNet backbone network structure. (Blue blocks are down-sampled to get orange blocks, and orange blocks are down-sampled to get purple blocks). The re-identification network uses fine-grained branches in the entire network shown in Figure 7 through HRNet (in Figure 7 branch3) for attitude estimation, so that the features of pedestrians are divided into 9 fine-grained features.

are carried out repeatedly to accomplish multi-scale information fusion.

E. RE-IDENTIFICATION NETWORK

In order to avoid the detailed information of the samples being ignored, this paper uses a modified network of the MGN for re-identification is shown in Figure 7, using ResNet-50 as the base network part. In terms of semantic-level features, three independent branches are included. The first branch enables capturing and acquiring general and global information about the image. In contrast, the second branch mainly divides the image into different parts to obtain the corresponding coarse-grained semantic data and improve the performance through the ECA attention mechanism. In the third branch, 17 key points are found by estimating the pose of the original image through the HRNet network, and these key points are divided into nine parts by semantics, as shown in Figure 4. Each branch collects and acquires information separately to learn more about pedestrians. Different branches have different tasks, yet they cooperate, and the first three lower layers are mainly shared in terms of weight. In comparison, the subsequent higher layers are relatively more independent and flexible regarding weights, which can effectively present general and local information.

Softmax loss is used for classification, and triple loss is used for metric learning to release the discriminative power of the learned representation of this network structure. Regarding discriminative learning, this topic is uniformly treated as equivalent to a multi-class classification problem when dealing with recognition tasks. Thus when dealing with the ith learning feature f_i, the Softmax loss formulation is explicitly shown as follows.

$$L_{soft\max} = -\sum_{i=1}^{N} \log \frac{e^{W_{y_i}^T f_i}}{\sum_{k=1}^{C} e^{W_k^T f_i}} \tag{3}$$

where W_k corresponds to the weight vector of the k class, the small batch size in the training process N and the number of classes in the training dataset C. Unlike the traditional Softmax loss, the application of the bias term provided within the prior linear multiclass classifier is discarded during, prompting a significant improvement in recognition performance. Softmax loss is used for a subset of features after normalization $\left\{f_{p_i}^{P_2}\,|_{i=1}^{2}\,, f_{p_i}^{P_3}\,|_{i=1}^{3}\right\}$ in all learned embeddings.

Triple loss training is performed on all global features of $\left\{f_g^G, f_g^{P_2}, f_g^{P_3}\right\}$ after the reduction to improve the ranking performance. The formula of the Triplet loss function used in this paper is as follows.

$$L_{triplet} = -\sum_{i=1}^{P}\sum_{a=1}^{K}\left[\begin{array}{c}\alpha + \max\limits_{p=1\cdots K}\left\|f_a^{(i)} - f_p^{(i)}\right\|_2 \\ -\min\limits_{\substack{n=1\cdots K \\ j=1\cdots P \\ j\neq i}}\left\|f_a^{(i)} - f_n^{(i)}\right\|_2\end{array}\right]_+ \tag{4}$$

where $f_a^{(i)}, f_p^{(i)}$, and $f_n^{(i)}$ are the features extracted from the anchor sample, positive sample, and negative sample. where 1,2 and 3 are the features extracted from the anchor, positive, and negative samples. They are margin hyperparameters that control the difference between internal and intradistance. The positive and negative samples involved in the period represent pedestrians who embody the same or different identities from the anchor, respectively.

III. EXPERIMENTS AND RESULTS

A. DATASET

The data analysis experiments conducted in this paper are based on the Market-1501 and DukeMTMC-reID pedestrian datasets, which are extensive. The Market-1501 dataset is derived from publicly available information from Tsinghua University in 2015, which includes 1,501 pedestrian images captured by 26 cameras, of which the number of annotated pedestrian rectangular frame images is 32668. The training set incorporates 12936 images of 751 pedestrians from this information; the test set selects 19732 images left by 750 people. The dataset comes from publicly available information from Duke University in 2016, which includes 36411 pedestrian rectangular frames left by 1404 pedestrians captured by eight cameras. From the training set, 16522 images from the image information of 702 people were selected; from the test set, 19889 images left by 702 people were selected. The specific information is shown in Table 1.

TABLE 1. Details of the dataset.

Dataset	Training set		Test set	
	ID	Images	ID	Images
Market-1501	751	12936	750	19732
DukeMTMC-reID	702	16522	702	19889

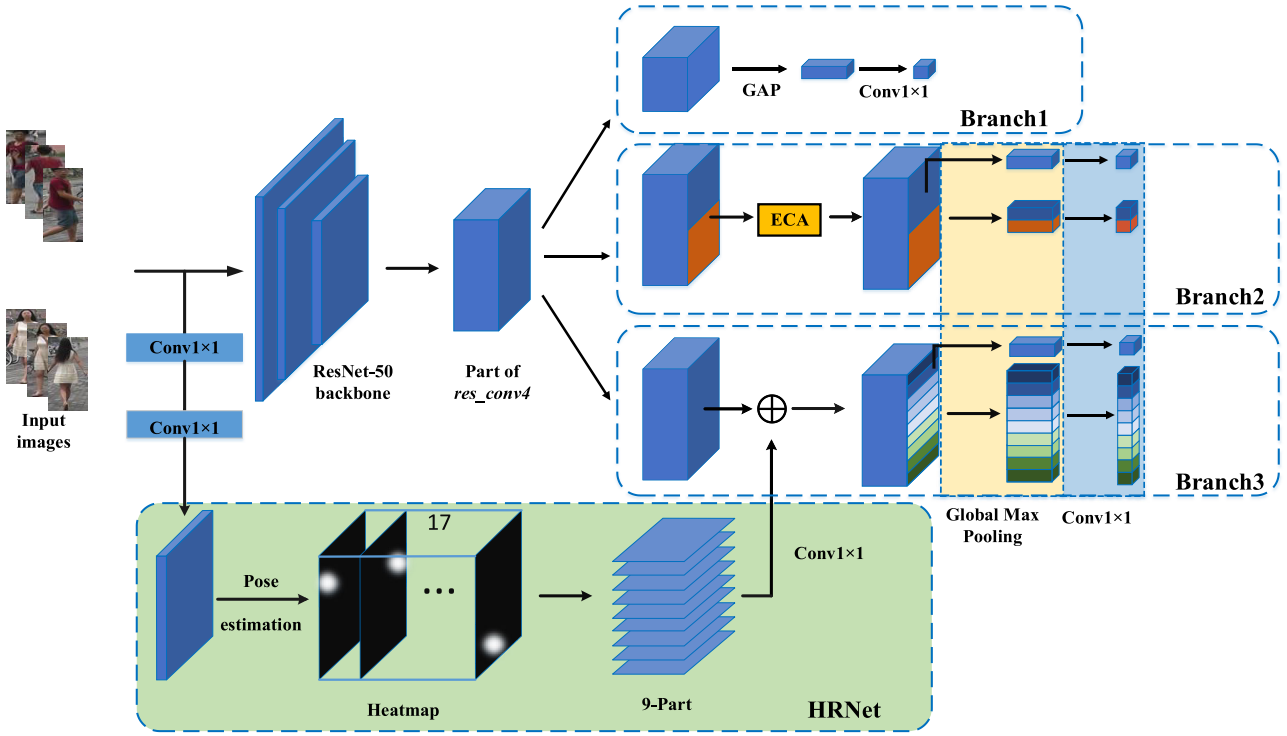

FIGURE 7. The overall structure of the re-identification network proposed in this paper. It consists of HRNet, MGN and ECA attention modules.

B. EXPERIMENTAL ENVIRONMENT

The experiments were conducted using Python 3.7 environment built with Anaconda, and Pytorch 1.8 was used for the deep learning framework.

The training of AOT-GAN was carried out sequentially Softmax loss is used for classification, and triple loss is used for metric learning to release the discriminative power of the learned representation of this network structure. Regarding discriminative learning, on Market-1501 and DukeMTMC-reID datasets. 12936 and 16522 non-masked images were applied for random masking during this period. The relevant details are shown: a rectangular block of the same size (i.e., 20 × 20) is generated at a random position of the non-occluded training image assigned to 255, thus obtaining the desired occluded image and the corresponding occluded and non-occluded image pairs. Since the image resolution sizes of the two datasets are different, the input image sizes are set to 64 × 128, and other optimization strategies for the generative adversarial network are developed with the Adam optimizer function in play. The generator and discriminator's learning rate is 1e-4, the Batchsize is 4, and Epoch is 1e5.

The input image size of the improved MGN network is still set to 64 × 128, and the optimization of the network is implemented using the Adam optimizer. Its learning rate base value is 2e-4, and Batchsize is 8.

C. EXPERIMENTAL RESULTS

First, the random occlusion map within the training set of Market1501 and DukeMTMC-reID is unfolded to remove the occlusion process, and then the corresponding generated image is obtained, as shown in Figure 8.

FIGURE 8. The pedestrian images in the Market1501 and DukeMTMC-reID training sets are randomly occluded and restored. (each group of images is the original image, the occluded image, and the restored image from left to right).

Combining the contents of Fig. 8, we can find that the series of pedestrian images distributed in the training set are randomly masked and then further de-masked based on the generator to obtain different pedestrian images from the original images. The left set of images is the Market1501 dataset, and the right set of images is the DukeMTMC-reID

FIGURE 9. Visualization of results. (The first row is the Market1501 dataset, and the second row is the DukeMTMC-reID dataset.)

TABLE 2. Comparison of the performance and completion time (s)of this paper on the Market-1501 dataset with mainstream algorithms.

Method	TIME(s)	RANK-1(%)	mAP(%)
SPREID[22]	5~10s	86.4	84.7
PCB+RPP[23]	10~15s	86.8	88.6
HPM[24]	15~20s	89.1	86.0
BFE[25]	15~20s	89.8	89.8
DCDS[26]	20~30s	90.1	87.3
IANET[27]	20~30s	90.4	88.0
GRL[28]	25~30s	91.3	89.1
HG[29]	25~30s	91.8	89.8
BPBREID[30]	35~40s	92.3	90.3
BIC-NET[31]	35~40s	93.0	91.2
OURS	25~30s	**93.2**	**91.6**

dataset, from left to right, the original, randomly masked, and de-masked images. The generated unmasked images are then re-identified by the reID network, and the example image is shown in Figure 9. The top image is based on the Market-1501 dataset, and the bottom image is based on the Duke-MTMC-reID dataset.

In order to ensure the outstanding feasibility and superiority of the designed scheme, this paper compares the advanced ReID algorithms SPREID, PCB+RPP, HPM, BFE, GRL, HG, BPBREID, BIC-NET and other series of methods with objective comparison, which revolves around the recognition of different strategies on the original and random occlusion maps. The cor-responding comparison results can be referred to in Table 2 and Table 3. This means that the algorithm's performance in this paper is outstanding and has high practical value.

D. ABLATION EXPERIMENTS

In order to verify the feasibility of the proposed multi-granularity generative adversarial network combined with the random occlusion method, ECA attention mechanism, and HRNet pose estimation method, a series of ablation experiments are designed in this paper using the Market1501 dataset and DukeMTMC-reID dataset. The

TABLE 3. Comparison of the performance and completion time (s)of this paper on the DukeMTMC-reID dataset with mainstream algorithms.

Method	TIME(s)	RANK-1(%)	mAP(%)
SPREID[16]	5~10s	83.3	80.9
PCB+RPP[17]	10~15s	83.6	81.6
HPM[18]	15~20s	84.9	82.3
BFE[19]	15~20s	85.5	83.8
DCDS[20]	20~30s	86.2	84.3
IANET[21]	20~30s	86.7	85.0
GRL[28]	25~30s	88.5	86.1
HG[29]	25~30s	88.9	86.5
BPBREID[30]	35~40s	89.5	87.3
BIC-NET[31]	35~40s	90.3	88.0
OURS	25~30s	**90.8**	**88.3**

TABLE 4. Performance comparison of ECA attention added to different branches of the network.

ECA Location	Market1501		DukeMTMC-reID	
	RANK-1(%)	mAP(%)	RANK-1(%)	mAP(%)
None	90.5	88.3	84.6	80.8
Branch1	92.1	89.0	86.8	82.4
Branch2	**93.2**	**91.6**	**90.8**	**88.3**
Branch3	92.8	90.5	89.2	87.4

TABLE 5. Comparison of the performance of different models.

Model	Market1501		DukeMTMC-reID	
	RANK-1(%)	mAP(%)	RANK-1(%)	mAP(%)
MGN	90.5	88.3	84.6	80.8
MGN+ECA	91.3	87.4	84.9	82.0
MGN+HRNet	92.6	90.8	89.2	86.1
MGN+ECA+HRNet	**93.2**	**91.6**	**90.8**	**88.3**

performance impact of the ECA attention mechanism added on different branches of the network on ReID is shown in Table 4.

The feasibility of the improvements for the MGN network, i.e., the addition of the ECA attention mechanism in Branch2 and the further segmentation of Branch3 using HRNet pose estimation at fine granularity for performance improvement, is shown in Table 5.

IV. CONCLUSION

In this paper, we design an algorithm for the re-recognition occluded pedestrians based on multi-grain generative adversarial networks. The randomly blocked pedestrian images are recovered using the generative adversarial network, and the retrieved images are passed into the improved multi-granularity network in this paper for re-recognition. The ability of the network to pay attention to coarse-grained features is improved by adding the ECA attention module to the coarse-grained branch of the multi-grained network. The multi-grained network's ability to pay attention to coarse-grained features is improved by adding the ECA attention module to the coarse-grained branch, and the fine-grained branch is posed estimation through the HRNet network to the pedestrian picture in 9 parts so that the network can obtain more pedestrian information and enhance its capacity to re-identify people.

The approach is tested and evaluated, and the associated experimental design is finished with satisfying results on various person-identification datasets reflecting harsh conditions to ensure the objectivity and viability of the analysis strategy presented in this study.

REFERENCES

[1] H. Luo, W. Jiang, and X. Fan, "Research progress on pedestrian re-identification based on deep learning," *J. Automat.*, vol. 45, no. 11, pp. 2032–2049, 2019.

[2] W. Li and X. Wang, "Locally aligned feature transforms across views," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3594–3601.

[3] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 886–893.

[4] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Oct. 1999, pp. 1150–1157.

[5] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2360–2367.

[6] N. Gheissari, T. B. Sebastian, and R. Hartley, "Person reidentification using spatiotemporal appearance," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2006, pp. 1528–1535.

[7] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2008, pp. 262–275.

[8] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3586–3593.

[9] Y. Ge, Z. Li, H. Zhao, G. Yin, S. Yi, and X. Wang, "FD-GAN: Pose-guided feature distilling GAN for robust person re-identification," in *Proc. Conf. Neural Inf. Process. Syst. (NIPS)*, 2018, pp. 1–15.

[10] H. Fan, L. Zheng, C. Yan, and Y. Yang, "Unsupervised person re-identification: Clustering and fine-tuning," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 14, no. 4, pp. 1–18, Nov. 2018.

[11] G. Wang, T. Zhang, J. Cheng, S. Liu, Y. Yang, and Z. Hou, "RGB-infrared cross-modality person re-identification via joint pixel and feature alignment," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3622–3631.

[12] S. Kanwal, J. H. Shah, M. A. Khan, M. Nisa, S. Kadry, M. Sharif, M. Yasmin, and M. Maheswari, "Person re-identification using adversarial haze attack and defense: A deep learning framework," *Comput. Electr. Eng.*, vol. 96, Dec. 2021, Art. no. 107542.

[13] J. Pang, D. Zhang, H. Li, W. Liu, and Z. Yu, "Hazy re-ID: An interference suppression model for domain adaptation person re-identification under inclement weather condition," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2021, pp. 1–6.

[14] X. Zhu, K. Guo, S. Ren, B. Hu, M. Hu, and H. Fang, "Lightweight image super-resolution with expectation-maximization attention mechanism," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1273–1284, Mar. 2022.

[15] X. Zhu, K. Guo, H. Fang, L. Chen, S. Ren, and B. Hu, "Cross view capture for stereo image super-resolution," *IEEE Trans. Multimedia*, vol. 24, pp. 3074–3086, 2022.

[16] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou, "Learning discriminative features with multiple granularities for person re-identification," in *Proc. 26th ACM Int. Conf. Multimedia*, Oct. 2018, pp. 274–282.

[17] Y. Zeng, J. Fu, H. Chao, and B. Guo, "Aggregated contextual transformations for high-resolution image inpainting," *IEEE Trans. Vis. Comput. Graph.*, vol. 29, no. 7, pp. 3266–3280, Jul. 2023.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[19] J. Li, B. M. Chen, and G. H. Lee, "SO-Net: Self-organizing network for point cloud analysis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9397–9406.

[20] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-Net and its variants for medical image segmentation: A review of theory and applications," *IEEE Access*, vol. 9, pp. 82031–82057, 2021.

[21] S. Seong and J. Choi, "Semantic segmentation of urban buildings using a high-resolution network (HRNet) with channel and spatial attention gates," *Remote Sens.*, vol. 13, no. 16, p. 3087, Aug. 2021.

[22] M. M. Kalayeh, E. Basaran, M. Gokmen, M. E. Kamasak, and M. Shah, "Human semantic parsing for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1062–1071.

[23] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, "Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)," in *Proc. ECCV*, 2018, pp. 480–496.

[24] Y. Fu, Y. Wei, Y. Zhou, H. Shi, G. Huang, X. Wang, Z. Yao, and T. Huang, "Horizontal pyramid matching for person re-identification," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 8295–8302.

[25] Z. Dai, M. Chen, X. Gu, S. Zhu, and P. Tan, "Batch DropBlock network for person re-identification and beyond," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3690–3700.

[26] L. T. Alemu, M. Shah, and M. Pelillo, "Deep constrained dominant sets for person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9854–9863.

[27] R. Hou, B. Ma, H. Chang, X. Gu, S. Shan, and X. Chen, "Interaction- and aggregatin network for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 9317–9326.

[28] X. Lin, P. Ren, C.-H. Yeh, L. Yao, A. Song, and X. Chang, "Unsupervised person re-identification: A systematic survey of challenges and solutions," 2021, *arXiv:2109.06057*.

[29] A. Rahimpour and H. Qi, "Attention-based few-shot person re-identification using meta learning," 2018, *arXiv:1806.09613*.

[30] V. Somers, C. D. Vleeschouwer, and A. Alahi, "Body part-based representation learning for occluded person re-identification," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 1613–1623.

[31] K. Zhou, Y. Yang, A. Cavallaro, and T. Xiang, "Learning generalisable omni-scale representations for person re-identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 5056–5069, Sep. 2022.

YANQI WANG was born in Chaoyang, Liaoning, China, in May 1999. He is currently pursuing the master's degree with Dalian Polytechnic University, Dalian, China. His current research interests include object detection and person re-identification.

YANGUO SUN was born in April 1976. Currently, he is the Director of the Information Center of the Second Hospital, Dalian Medical University, where he has been engaged in hospital information construction for more than ten years. He is also an associate professor and a system analyst. He has presided over and participated in standardization and informatization construction projects, such as grade hospital evaluation. His current research interest includes medical image processing. He is a young member of the 8th Committee of the Chinese Medical Informatics Branch and a member of the Telemedicine Informatization Professional Committee of the China Health Information and Healthcare Big Data Society and the Health Card Application Management Professional Committee of the China Health Information and Healthcare Big Data Society. He won the Third Prize in the Hospital Science and Technology Innovation Award of the China Hospital Association. He has chaired and participated in several provincial and municipal scientific research projects.

ZHENPING LAN was born in September 1978. She is currently an Assistant Professor with the Communication Engineering Department, School of Information Science and Engineering, Dalian Polytechnic University, Dalian, China. She has published more than 30 journals. She has presided and participated in more than 20 vertical and horizontal projects, with a total of more than two million, one invention patent, and two provincial third prizes for scientific and technological achievements. Her current research interests include deep learning and mobile communication.

FENGXUE SUN was born in Changchun, Jilin, China. She is currently pursuing the master's degree with Dalian Polytechnic University, Dalian, China. Her current research interest includes AI image processing technology.

NIANCHAO ZHANG was born in Weifang, Shandong, China, in May 1999. He is currently pursuing the master's degree with Dalian Polytechnic University, Dalian, China. His current research interest includes deep learning.

YURU WANG received the M.S. degree in engineering. Currently, she is a Lecturer with the Automation Department, School of Information Science and Engineering, Dalian Polytechnic University, Dalian, China. Her current research interests include embedded applications and intelligent control.